

Db 181 GGATGTGGCTTGTACTCTAGCCAAATGCCAGGGTACTTCAAGAGTACCAGAACT 240
Qy 241 CAAGGAGGGGTCAAGCAGCTGA 263
Db 241 CAAGGAGGGGTCAAGCAGCTGA 263
RESULT 2
AF303227 175 bp DNA linear BCT 21-MAR-2001
LOCUS Streptococcus suis clone ivs 31 in vivo selected promoter sequence.
DEFINITION AF303227
ACCESSION AF303227.1 GI:13398370
VERSION AF303227.1
KEYWORDS Streptococcus suis
SOURCE Streptococcus suis
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 175)
AUTHORS Smith,H.E., Buijs,H., de Vries R.R., Wisselink,H.J.,
Stockhofe-Zurwieden,N. and Smits,M.A.
TITLE Environmentally regulated genes of Streptococcus suis:
identification by the use of iron-restricted conditions in vitro
and by experimental infection of piglets
JOURNAL Microbiology 147 (Pt 2), 271-280 (2001)
PUBMED 11158344
REFERENCE 2 (bases 1 to 175)
AUTHORS Smith,H.E., Buijs,H., de Vries,R., Wisselink,H.J.,
Stockhofe-Zurwieden,N. and Smits,M.A.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Department of Bacteriology, Institute for
Animal Science and Health, P.O.Box 65, Lelystad 8200 AB, The
Netherlands
FEATURES
source 1. .175
misc_feature 1. .175
note="in vivo selected promoter sequence"
ORIGIN
Query Match 66.2%; Score 174; DB 1; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.4e-43;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 89 CTCCTGACCACCTATNTGCATCAAGTCCAAATGACAGTCGAGTGTGCGGTAGACAAC 148
Db 1 CTCCTGACCACCTATNTGCATCAAGTCCAAATGACAGTCGAGTGTGCGGTAGACAAC 60
Qy 149 TACTATACGGCAGGAAGTGGAGATTGAGTTGGATTGGCTTTGACTCTCTAGCCAAAAT 208
Db 61 TACTATACGGCAGGAAGTGGAGATTGAGTTGGATTGGCTTTGACTCTCTAGCCAAAAT 120
Qy 209 GCCCAGCGGTACTTCAAGAAGTACAGAAACTCAAGGAGGCGGTCAAGCAGCTGA 263
Db 121 GCCCAGCGGTACTTCAAGAAGTACAGAAACTCAAGGAGGCGGTCAAGCAGCTGA 175
RESULT 3
AY566303 1938 bp DNA linear BCT 23-MAR-2004
LOCUS Streptococcus suis strain 9801 FbpS (fbps) gene, complete cds.
DEFINITION AY566303
ACCESSION AY566303
VERSION AY566303.1 GI:45549555
KEYWORDS Streptococcus suis
SOURCE Streptococcus suis
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 1938)
AUTHORS

AUTHORS Sun,L., Fan,H. and Lu,C.
TITLE Cloning and Sequence Analysis of FbpS Gene of Streptococcus suis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1938)
AUTHORS Sun,L., Fan,H. and Lu,C.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-2004) College of Veterinary Medicine, Nanjing
Agricultural University, No. 1 Tongwei Road, Nanjing, Jiangsu
210095, P.R. China
FEATURES
source 1. .1938
Location/Qualifiers
/organism="Streptococcus suis"
/mol_type="genomic DNA"
/strain="9801"
/serotype="2"
/specific host="swine"
/db_xref="taxon:1307"
/country="China"
83. .1741
/gene="fbpS"
83. .1741
/gene="fbpS"
/note="fn/fgm-binding protein"
/codon_start=1
/transl_table=11
/product="FbpS"
/protein_id="AA567693.1"
/db_xref="GI:45549555"
/translation="MSPDGFLLHMTAELRANLEGGRIQKINQPFQEIIVLINIRNQ
SHKLLSAHSVFGRVQLTQSDFTNPKVPNTFTMLRKYLOGAIEEIROLDNDILEF
SVNKDIGHIQAATLIVEIMKHSNLLVDKSEQKIIIEAKHVPSONSVRTILPGS
TYIRPPTHSLNPTVSEDKFELLSTQELSPKQLQVFGQLGRDTASELANHQLQIDR
LKNFAFFDQATPSLTDKSYAALPFANSPENOPHFESLSLDDFFYQDKAERDVAQ
QANELIKRVAEELKREKLIKQEGILADTETAEIVKQKGLLTYLHQVNDQSSVR
LDNYTGKLEIELDVALTPSONAQRVFKYOKLKEAVKHLNIEETKSTIVLESV
DTMLQASLAEIDIEELIETGYLKERHEKIHKROKPERVLTADGKTIILVGNL
QNDELTFMAKKGELWFHAKDIPGSHVIVTDNLDPSEVTKDAELAAAYFSKARHNL
VQVDMIEAKKLHPTGKPGFVYVRGKTLKRVTPTEDKIKSMKI"
ORIGIN
Query Match 66.2%; Score 174; DB 1; Length 1938;
Best Local Similarity 99.4%; Pred. No. 1.5e-43;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 89 CTCCTGACCACCTATNTGCATCAAGTCCAAATGACAGTCGAGTGTGCGGTAGACAAC 148
Db 1034 CTCCTGACCACCTATNTGCATCAAGTCCAAATGACAGTCGAGTGTGCGGTAGACAAC 1093
Qy 149 TACTATACGGCAGGAAGTGGAGATTGAGTTGGATTGGCTTTGACTCTCTAGCCAAAAT 208
Db 1094 TACTATACGGCAGGAAGTGGAGATTGAGTTGGATTGGCTTTGACTCTCTAGCCAAAAT 1153
Qy 209 GCCCAGCGGTACTTCAAGAAGTACAGAAACTCAAGGAGGCGGTCAAGCAGCTGA 263
Db 1154 GCCCAGCGGTACTTCAAGAAGTACAGAAACTCAAGGAGGCGGTCAAGCAGCTGA 1208
RESULT 4
AF438158 2179 bp DNA linear BCT 05-MAR-2002
LOCUS Streptococcus suis fibronectin/fibrinogen binding protein (fbps)
DEFINITION gene, complete cds; and alpha-acetolactate decarboxylase gene,
partial cds.
ACCESSION AF438158
VERSION AF438158.1 GI:19110785
KEYWORDS Streptococcus suis
SOURCE Streptococcus suis
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 2179)
AUTHORS de Greeff,A., Buys,H., Verhaar,R., Dijkstra,J., van Alphen,L. and

Smith,H.E.
Contribution of fibronectin-binding protein to pathogenesis of
Streptococcus suis serotype 2
Infect. Immun. 70 (3), 1319-1325 (2002)
11854216
2 (bases 1 to 2179)
de Greeff,A., Buys,H., Verhaar,R., Dijkstra,J., van Alphen,L. and
Smith,H.
Direct Submission
Submitted (22-OCT-2001) Infectious Diseases and Food Chain Quality,
ID-Lelystad, Deelherweg 15, Lelystad NL-8219PH, The Netherlands
Location/Qualifiers
1. .2179
/organism="Streptococcus suis"
/mol_type="genomic DNA"
/db_xref="taxon:1307"
121. .129
/rpt_type=inverted
136. .1905
/gene="fbps"
136. .141
/genes="fbps"
143. .151
/rpt_type=inverted
159. .164
/gene="fbps"
184. .190
/genes="fbps"
207. .212
/genes="fbps"
235. .241
/gene="fbps"
247. .1905
/genes="fbps"
/note="FBPS"
/codon_start=1
/transl_table=11
/product="fibronectin/fibrinogen binding protein"
/protein_id="AAU85276.1"
/db_xref="GI:19110786"
/translation="MSFDGDFLLHMTAELRANLEGGRIOKIHOPFBEIVLNIQRNQ
SHKLLSAHSVFRVOLQSDFTNPVNTFTMLRKYLQGAIEIRIQDNDRILEF
SVNKGIDGHIQATLIVEIMGKHSNIILVDKSEOKIEAIKHVGFSPQSYRTILPGS
TWIRPETHSLNPYVDEKLEILSTOELSPKNLOQVFGLRDTASLANHLQIDR
LKNFAPFOATOPSITDKSYAALPANSFENOPHFESLSLLDFFYQKASRDRVAQ
QANELLKRVASELEKRNKLLIKQELADTETAEALVRQKGLLTTTHVQVPNDQSVR
LDNYTKLELIEIDVALTPSQAQRYFKYQKLEAVKHLNLIIEETKSTIVYLESV
DTMLQASLAEIDIEIREELIETGLKRRHREKIKRQKPERYLATDGTCTIILVGNL
ONDELTFKMAKGELEWPHAKDIIPGSHVITDNLDPDVKTDAAELAAVFSKARHSNL
VQVDMIEAKLHKPTGKQGFVYRGQKTLRVPTEDKIKSMKI"
complement(1931. .>2179)
/codon_start=1
/transl_table=11
/product="alpha-acetolactate decarboxylase"
/protein_id="AAU85277.1"
/db_xref="GI:19110787"
/translation="GTIVGIWTEIFHGVSVAGYHLHFTSDLLTFGHWMDYVISEGM
VEVGPVDQLDQRPVQDQRYLYAKFNAKEVREDIDKAB"
ORIGIN
Query Match 66.2%; Score 174; DB 1; Length 2179;
Best Local Similarity 99.4%; Pred. No. 1.5e-43;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 89 CTCCTGACCACTTATGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 148
DB 1198 CTCTGACCACTTATGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 1257
QY 149 TACTATACGGCGCAAGCACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAAC 208
DB 1258 TACTATACGGCGCAAGCACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAAC 1317
QY 209 GCCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAACCTGA 263

Db 1318 GCCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAACCTGA 1372
RESULT 5
CP000023_08/c
WPCOMMENT
Sequence split into 18 fragments LOCUS CP000023 Accession CP000023
Fragment Name Begin End
CP000023_00 1 110000
CP000023_01 100001 210000
CP000023_02 200001 310000
CP000023_03 300001 410000
CP000023_04 400001 510000
CP000023_05 500001 610000
CP000023_06 600001 710000
CP000023_07 700001 810000
CP000023_08 800001 910000
CP000023_09 900001 1010000
CP000023_10 1000001 1110000
CP000023_11 1100001 1210000
CP000023_12 1200001 1310000
CP000023_13 1300001 1410000
CP000023_14 1400001 1510000
CP000023_15 1500001 1610000
CP000023_16 1600001 1710000
CP000023_17 1700001 1796846
Continuation (9 of 18) of CP000023 from base 800001 (CP000023 Streptococcus thermophilus
Query Match 35.7%; Score 94; DB 1; Length 110000;
Best Local Similarity 70.9%; Pred. No. 3.2e-18;
Matches 124; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 89 CTCCTGACCACTTATGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 148
DB 5778 CTTTAAACAACCTTCTGAGCATGTGTGCGCAAAATGACAGATAGTGTAGACTGACCAAC 57719
QY 149 TACTATACGGCGCAAGCACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAAC 208
DB 57718 TACTATACGGCGCAAGCACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAAC 57659
QY 209 GCCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAACCTGA 263
DB 57658 GCTCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGCGGTCGAGCAACCTGA 57604
RESULT 6
CP000024_08/c
WPCOMMENT
Sequence split into 18 fragments LOCUS CP000024 Accession CP000024
Fragment Name Begin End
CP000024_00 1 110000
CP000024_01 100001 210000
CP000024_02 200001 310000
CP000024_03 300001 410000
CP000024_04 400001 510000
CP000024_05 500001 610000
CP000024_06 600001 710000
CP000024_07 700001 810000
CP000024_08 800001 910000
CP000024_09 900001 1010000
CP000024_10 1000001 1110000
CP000024_11 1100001 1210000
CP000024_12 1200001 1310000
CP000024_13 1300001 1410000
CP000024_14 1400001 1510000
CP000024_15 1500001 1610000
CP000024_16 1600001 1710000
CP000024_17 1700001 1796226
Continuation (9 of 18) of CP000024 from base 800001 (CP000024 Streptococcus thermophilus
Query Match 35.7%; Score 94; DB 1; Length 110000;
Best Local Similarity 70.9%; Pred. No. 3.2e-18;


```
ACCESSION AR653221
VERSION AR653221.1 GI:67583508
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillion, P.J., Dougherty, B.,
Fannon, M.R., and Rosen, C.A.
TITLE Streptococcus pneumoniae SP036 polynucleotides
JOURNAL Patent: US 6887663-A 59 03-MAY-2005;
Human Genome Sciences, Inc.; Rockville, MD
FEATURES
source
Location/Qualifiers
1..1423
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1423;
Best Local Similarity 67.8%; Pred. No. 1.9e-15;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACTATNTGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 149
Db 717 TGCTGACCACTTCCTCCACCAAGTGCCTAACGACCAAGACCAAGGTTATCTCTAGACAAC 776
QY 150 ACTATACGGCGCAAGCAAGTGGAGATTGAGTTGGATGTGGCTTTGACTCTCTAGCAAAATG 209
Db 777 ACTATACCAACCAACCTATCATGATTGCGCTTGATTAAGGCTCTGACTCCCAACCAAGATG 836
QY 210 CCCAGCGGTACTTCAAGAGTAGTACCAAACTCAAGGAGCGGTCAAGCACTGA 263
Db 837 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 890
RESULT 11
AX568696 1680 bp DNA linear PAT 29-NOV-2002
LOCUS
DEFINITION Sequence 1903 from Patent WO02077021.
ACCESSION AX568696
VERSION AX568696.1 GI:26002414
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Masignani, V., Tettelin, H., and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 1903 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
Location/Qualifiers
1..1680
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"
ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1680;
Best Local Similarity 67.8%; Pred. No. 1.9e-15;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACTATNTGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 149
Db 974 TGCTGACCACTTCCTCCACCAAGTGCCTAACGACCAAGACCAAGGTTATCTCTAGACAAC 1033
QY 150 ACTATACGGCGCAAGCAAGTGGAGATTGAGTTGGATGTGGCTTTGACTCTCTAGCAAAATG 209
Db 1034 ACTATACCAACCAACCTATCATGATTGCGCTTGATTAAGGCTCTGACTCCCAACCAAGATG 1093
QY 210 CCCAGCGGTACTTCAAGAGTAGTACCAAACTCAAGGAGCGGTCAAGCACTGA 263
Db 1094 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1147
```

```
RESULT 12
AR481615 1683 bp DNA linear PAT 14-MAY-2004
LOCUS
DEFINITION Sequence 2519 from patent US 6699703.
ACCESSION AR481615
VERSION AR481615.1 GI:47240577
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1683)
AUTHORS Doucette-Stamm, L., Bush, D., Zeng, Q., Opperman, T. and
Houseweart, C.E.
TITLE Nucleic acid and amino acid sequences relating to Streptococcus
pneumoniae for diagnostics and therapeutics
JOURNAL Patent: US 6699703-A 2519 02-MAR-2004;
Genome Therapeutics Corporation; Waltham, MA
FEATURES
source
Location/Qualifiers
1..1683
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1683;
Best Local Similarity 67.8%; Pred. No. 1.9e-15;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACTATNTGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 149
Db 974 TGCTGACCACTTCCTCCACCAAGTGCCTAACGACCAAGACCAAGGTTATCTCTAGACAAC 1033
QY 150 ACTATACGGCGCAAGCAAGTGGAGATTGAGTTGGATGTGGCTTTGACTCTCTAGCAAAATG 209
Db 1034 ACTATACCAACCAACCTATCATGATTGCGCTTGATTAAGGCTCTGACTCCCAACCAAGATG 1093
QY 210 CCCAGCGGTACTTCAAGAGTAGTACCAAACTCAAGGAGCGGTCAAGCACTGA 263
Db 1094 CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1147
RESULT 13
AR587857 1692 bp DNA linear PAT 15-DEC-2004
LOCUS
DEFINITION Sequence 1976 from patent US 6800744.
ACCESSION AR587857
VERSION AR587857.1 GI:56632114
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1692)
AUTHORS Doucette-Stamm, L.A. and Bush, D.
TITLE Nucleic acid and amino acid sequences relating to Streptococcus
pneumoniae for diagnostics and therapeutics
JOURNAL Patent: US 6800744-A 1976 05-OCT-2004;
Genome Therapeutics Corporation; Waltham, MA
FEATURES
source
Location/Qualifiers
1..1692
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 32.3%; Score 85; DB 6; Length 1692;
Best Local Similarity 67.8%; Pred. No. 1.9e-15;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACTATNTGTCATCAAGTCCCAATGACCAAGTCGAGTGTGCGGTAGACAAC 149
Db 983 TGCTGACCACTTCCTCCACCAAGTGCCTAACGACCAAGACCAAGGTTATCTCTAGACAAC 1042
QY 150 ACTATACGGCGCAAGCAAGTGGAGATTGAGTTGGATGTGGCTTTGACTCTCTAGCAAAATG 209
```

Db 1043 ACTATACCAACCACTATCATGATTGGCTTGATAAGGCTCTGACTCCCAACAGATG 1102
 Qy 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGGCGGTCAAGCACCTGA 263
 Db 1103 CCCACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1156

RESULT 14
 AF181976 2015 bp DNA linear BCT 02-NOV-1999
 DEFINITION Streptococcus pneumoniae adherence and virulence protein A (pav)
 gene, complete cds.
 ACCESSION AF181976
 VERSION AF181976.1 GI:6175914
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE 1 (bases 1 to 2015)
 Holmes,A.R., McNab,R., Millsap,K. and Jenkinson,H.F.
 TITLE The pava gene of Streptococcus pneumoniae encodes a
 fibronectin-binding protein that is necessary for Pneumococcal cell
 adhesion and virulence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2015)
 Holmes,A.R., McNab,R. and Jenkinson,H.F.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-1999) Oral Sciences and Orthodontics, University
 of Otago, Great King St, Dunedin, New Zealand

FEATURES
 source
 1..2015
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /strain="R800"
 /db_xref="taxon:1313"
 98..1763
 /gene="pava"
 98..102
 /gene="pavA"
 108..1763
 /gene="pavA"
 /function="adhesin"
 /codon_start=1
 /transl_table=11
 /product="adherence and virulence protein A"
 /protein_id="AAF05332.1"
 /db_xref="GI:6175915"
 /translation="MSFPGFLHIVELRSELVNGRIQKINQPEQLVLQIRNRQ
 SHRLLSHPVFGRIQLTQTTPENPAQSTFIMVLKYLQGLALIESIQVENDRIVEI
 TVSNKEIGDIQATLIIEINGKSNILLVDKSHKILEVHKHVFQSONSYRTLLPGS
 TYIAPPSTNPFTIKDEKLFILQTELTAQNLSLQFGLGRDTANELERILVSEK
 LSAPFNPNQCTCLTFTSPVFNQAGEPFLANSLDLDTYKNAERDRVKQQA
 SELIRRVENELKQKHKLQERELLADNAEERQKGLLTTLHQPNDQDVIDL
 NYTNPQIMIALDKALTNQAQYFKKYQKLKAVKYLTLIETKATILYLSVET
 VINQGLBEIAEIRRELQTFIRRRQREKIQKRKUEQYLASDGKTIIVYGRNNLQN
 BELTFMKARKELMFHAKDIPGSHVIVISGNDPSDVKNTDAEELAAVPSQGRLSNLVQ
 VDMIEVKLNKPTGKPGFVTVTGQKTLRVTPDSPSKIASMKKS"

ORIGIN

Query Match 32.3%; Score 85; DB 1; Length 2015;
 Best Local Similarity 67.8%; Pred. No. 1.9e-15;
 Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 Qy 90 TCCTGACCACTATNTGATCAAGTCCAAATGACCACTCGAGTGTGGGTAGACAACT 149
 Db 1054 TGCTGACCACTTCTCCACCAAGTGCCTACGACCAAGACCAAGCTTATCTAGCAACT 1113
 Qy 150 ACTATACCGGCAAGAACTGGAGATTGAGTTGAGTGTGGCTTGTACTCTAGCCAAATG 209
 Db 1114 ACTATACCAACCACTATCATGATTGGCTTGATAAGGCTCTGACTCCCAACCAAGATG 1173

Qy 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGGCGGTCAAGCACCTGA 263
 Db 1174 CCCACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1227

RESULT 15
 CQ788941/c 10240 bp DNA linear PAT 29-MAR-2004
 LOCUS CQ788941
 DEFINITION Sequence 32 from Patent EP1400592.
 ACCESSION CQ788941
 VERSION CQ788941.1 GI:45822509
 KEYWORDS
 SOURCE unidentified
 ORGANISM unidentified
 unclassified sequences.

REFERENCE 1
 Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Barash,S.C.,
 Fannon,M. and Dougherty,B.A.
 TITLE Streptococcus pneumoniae polynucleotides and sequences
 JOURNAL Patent: EP 1400592-A 32 24-MAR-2004;
 HUMAN GENOME SCIENCES, INC. (US)
 FEATURES
 Location/Qualifiers
 1..10240
 /organism="unidentified"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32644"

ORIGIN
 Query Match 32.3%; Score 85; DB 6; Length 10240;
 Best Local Similarity 67.8%; Pred. No. 2e-15;
 Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 90 TCCTGACCACTATNTGATCAAGTCCAAATGACCACTCGAGTGTGGGTAGACAACT 149
 Db 4646 TGCTGACCACTTCTCCACCAAGTGCCTACGACCAAGACCAAGCTTATCTAGCAACT 4587
 Qy 150 ACTATACCGGCAAGAACTGGAGATTGAGTTGAGTGTGGCTTGTACTCTAGCCAAATG 209
 Db 4586 ACTATACCAACCACTATCATGATTGGCTTGATAAGGCTCTGACTCCCAACCAAGATG 4527
 Qy 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGGCGGTCAAGCACCTGA 263
 Db 4526 CCCACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 4473

Search completed: January 20, 2006, 15:56:27
 Job time : 2163 secs

Result No.	Score	Query			DB ID	Description
		Match	Length	DB		
1	260.4	99.0	263	6	ABS59824	AbS59824 Streptoco
2	85	32.3	1423	2	AAV27353	AAv27353 Streptoco
3	85	32.3	1423	6	ABQ84821	ABq84821 S. pneumo
4	85	32.3	1423	10	ADC45140	ADc45140 S. pneumo
5	85	32.3	1653	12	ADM91909	Adm91909 S. pneumon
6	85	32.3	1680	10	ABX06664	ABx06664 S. pneumo
7	85	32.3	1683	4	AA55569	AA55569 Streptoco
8	85	32.3	1683	8	ACA49876	ACa49876 Prokaryot
9	85	32.3	1683	13	ADK46004	ADk46004 Streptoco
10	85	32.3	1692	13	ADR93341	Adr93341 Novel. S.
11	85	32.3	1692	14	AA57211	AAa57211 Streptoco
12	85	32.3	10240	2	AAV52165	AAv52165 Streptoco
13	85	32.3	110000	10	ABS56454_07	Continuation (8 of
14	85	32.3	110000	10	ABS56454_08	Continuation (9 of
15	83.2	31.6	1653	8	ACA50476	ACa50476 Prokaryot
16	83.2	31.6	1728	6	ABN67397	ABn67397 Streptoco
17	76.2	29.0	1647	8	ACA48071	ACa48071 Prokaryot
18	66.8	25.4	1653	6	ABN67396	ABn67396 Streptoco
19	66.8	25.4	1656	13	ADV84865	ADv84865 Streptoco

sequence 1423 BP: 488 A: 324 C: 281 G: 330 T: 0 U: 0 Other;

Query Match 32.3%; Score 85; DB 10; Length 1423;

90 TCCTGACCACCTATNTGCATCAAGTGCCAAATGACCAGTCGAGTGTGCGGTAGACAACT 149

150 ACTATACGGCAAGGAAGGAGATTGAGTTGGATGTGGCTTTTGACTTCTAGCCAAATG 209

210 CCCAGCGTACTTCAAGAAGTACCAGAACTCAAGGAGCGGTCAAGCACCTGA 263

C ADM91909;

S pneumoniae antigenic protein-encoding gene sequence SeqID106.

X

X

R P-PSDB; ADM92146.

T pneumoniae.

This invention relates to novel isolated *Streptococcus pneumoniae* nucleic acids and the antigenic polypeptides encoded by them. The

compositions and methods disclosed are useful for treating streptococcus

CC pneumoniae infection. The present sequence is that of an S pneumoniae
 CC gene of the invention.
 XX
 SQ Sequence 1653 BP; 551 A; 383 C; 322 G; 397 T; 0 U; 0 Other;
 Query Match 32.3%; Score 85; DB 12; Length 1653;
 Best Local Similarity 67.8%; Pred. No. 3.6e-18;
 Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 QY 90 TCCTGACCACCTATNTGTCATCAAGTGCACCAATGACCACTCGAGTGTGGGTTAGACAACT 149
 DB 947 TGCTGACAACTTCTCCACCAAGTGCCTACGACCAAGACCAAGTTCCTAGACAACT 1006
 QY 150 ACTATACGGGCAAGAACTGGAGATTGATGTTGGATGCTTGAATCTCTAGCCAAATG 209
 DB 1007 ACTATACCAACCAACTATCATGATGTGGCTTGATGAGGCTCTGACTCCCAACCAAGATG 1066
 QY 210 CCCAGCGTACTTCAAGAAGTACCAAGAACTCAAGGCGGTCAGCACCTGA 263
 DB 1067 CCCAACGCTATTTTAAACGGTATCAGAAACTCAAGAAGCTGTCAATACTTGA 1120
 RESULT 6
 ABX06664
 ID ABX06664 standard; DNA; 1680 BP.
 XX
 AC ABX06664;
 XX
 DT 27-OCT-2003 (revised)
 DT 11-FEB-2003 (first entry)
 XX
 XX S. pneumoniae type 4 strain coding region #952.
 DE Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae; type 4 strain.
 XX
 FN WO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 XX 27-MAR-2002; 2002WO-IB002163.
 PF
 XX 27-MAR-2001; 2001GB-00007658.
 PR
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Masignani V, Tettelin H, Fraser C;
 FI
 XX WPI; 2003-040579/03.
 DR P-PSDB; ABU01377.
 XX
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
 PT ear infection.
 XX
 XX Claim 6; SEQ ID NO 1903; 56pp; English.
 PS
 XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS6454. Also included are an antibody which binds one of the proteins,
 CC treating a patient by administering the protein, DNA or antibody (in a
 CC composition), a kit comprising first and second primers, which are the
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a
 CC sequence not defined in the specification, for amplifying a target
 CC sequence contained within a Streptococcus nucleic acid sequence, where

CC the first primer is substantially complementary to the target sequence
 CC and the second primer is substantially complementary to the complement of
 CC the target sequence, and where the parts of the primers having
 CC substantial complementarity define the termini of the target sequence to
 CC be amplified, assay comprising contacting a test compound with the
 CC protein, and determining whether the test compound binds to the protein
 CC and a Streptococcus pneumoniae bacterium, where one or more genes
 CC encoding the proteins has been rendered inactive. The proteins, nucleic
 CC acid molecules, antibody and compositions are useful as medicaments for
 CC treating or preventing a disease or infection due to streptococcus
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
 CC media or ear infection. They are also useful in developing vaccines,
 CC diagnostics and antibiotics. The methods are useful for identifying
 CC immunodominant proteins. The present sequence is one of the 2489
 CC identified coding region from the genomic sequence. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 1680 BP; 564 A; 387 C; 329 G; 400 T; 0 U; 0 Other;
 Query Match 32.3%; Score 85; DB 10; Length 1680;
 Best Local Similarity 67.8%; Pred. No. 3.6e-18;
 Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 QY 90 TCCTGACCACCTATNTGTCATCAAGTGCACCAATGACCACTCGAGTGTGGGTTAGACAACT 149
 DB 974 TGCTGACAACTTCTCCACCAAGTGCCTACGACCAAGACCAAGTTCCTAGACAACT 1033
 QY 150 ACTATACGGGCAAGAACTGGAGATTGATGTTGGATGCTTGAATCTCTAGCCAAATG 209
 DB 1034 ACTATACCAACCAACTATCATGATGTGGCTTGATGAGGCTCTGACTCCCAACCAAGATG 1093
 QY 210 CCCAGCGTACTTCAAGAAGTACCAAGAACTCAAGGAGCGGTCAGCACCTGA 263
 DB 1094 CCCAACGCTATTTTAAACGGTATCAGAAACTCAAGAAGCTGTCAATACTTGA 1147
 RESULT 7
 AAS55569
 ID AAS55569 standard; DNA; 1683 BP.
 XX
 AC AAS55569;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 XX Streptococcus pneumoniae DNA for cellular proliferation protein #140.
 DE
 XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
 KW antibacterial; drug design.
 KW
 XX Streptococcus pneumoniae.
 OS
 XX WO200170955-A2.
 FN
 XX 27-SEP-2001.
 PD
 XX 21-MAR-2001; 2001WO-US009180.
 PF
 XX 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207272P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX

DR WPI; 2001-611495/70.
DR P-PSDB; AAU37710.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 9206; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes,
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1683 BP; 565 A; 387 C; 331 G; 400 T; 0 U; 0 Other;
Query Match 32.3%; Score 85; DB 4; Length 1683;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCTGACCACTATNTGATCAAGTCCAAATGACCAAGTCGAGTGGCGGTAGACAAC 149
Db 974 TGTGACAACTTCTCCACCAAGTGCCTTAACGACCAAGCAGGTTATCTTAGACAAC 1033
QY 150 ACTATACGGGCAAGGAACTGGAGATTGAGTTGGATGCTGCTTGAATCTCTAGCAAAATG 209
Db 1034 ACTATACCAACCAACCTATCATGTTGCGTTGATAAGGCTCTGACTCCCAACCAAGATG 1093
QY 210 CCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGGCGGTCAAGCACCTGA 263
Db 1094 CCACGCTATTTTAAACGGTATCAGAACTCAAGAGGCTCTCAATAACTTGA 1147
RESULT 8
ACA49876
ID ACA49876 standard; DNA; 1683 BP.
XX
AC ACA49876;
XX
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #31533.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Streptococcus pneumoniae.
XX
XX WO20027183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX

PA (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU46006.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 37746; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway; (8)
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1683 BP; 565 A; 387 C; 331 G; 400 T; 0 U; 0 Other;
Query Match 32.3%; Score 85; DB 8; Length 1683;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCTGACCACTATNTGATCAAGTCCAAATGACCAAGTCGAGTGGCGGTAGACAAC 149
Db 974 TGTGACAACTTCTCCACCAAGTGCCTTAACGACCAAGCAGGTTATCTTAGACAAC 1033
QY 150 ACTATACGGGCAAGGAACTGGAGATTGAGTTGGATGCTGCTTGAATCTCTAGCAAAATG 209
Db 1034 ACTATACCAACCAACCTATCATGTTGCGTTGATAAGGCTCTGACTCCCAACCAAGATG 1093
QY 210 CCAGCGGTACTTCAAGAGTACCAAGAACTCAAGAGGCGGTCAAGCACCTGA 263
Db 1094 CCACGCTATTTTAAACGGTATCAGAACTCAAGAGGCTCTCAATAACTTGA 1147
RESULT 9
ADK46004
ID ADK46004 standard; DNA; 1683 BP.
XX
XX ADK46004;
XX
XX 24-FEB-2005 (first entry)
XX

DE Streptococcus pneumoniae gene, Seq ID No 2519.
XX ds; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
KW Streptococcus pneumoniae.
XX US6699703-B1.
XX 02-MAR-2004.
XX 26-MAY-2000; 2000US-00583110.
XX 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CB;
XX WPI; 2004-212399/20.
DR P-PSDB; ADK48665.
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX Disclosure; SEQ ID NO 2519; 301pp; English.
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae nucleic acid of the invention. Note: The
CC sequence data for this patent did not appear in the printed specification
CC but was obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1683 BP; 563 A; 386 C; 332 G; 402 T; 0 U; 0 Other;
Query Match 32.3%; Score 85; DB 13; Length 1683;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
OY 90 TCCTGACCACCTATNTGTCATCAAGTGCACCAATGACACGTCGAGTGTGGGTTAGACAACT 149
DB 974 TGCTGACAACTTCTCTCCACCAAGTGCTTACGACCAAGACGAGTTATCTAGACAACT 1033
OY 150 ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTTAGCCAAATG 209
DB 1034 ACTATACCAACCAACCTATCATGATTGGCTTGTATAAGGCTCTGACTCCCAACCAAGATG 1093
OY 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACCTGA 263
DB 1094 CCCAACGCTATTTAAACGGTATCAGAACTCAAGAAAGCTGTCAATATTGA 1147
RESULT 10
AD93341
ID AD93341 standard; DNA; 1692 BP.
XX AC AD93341;
XX 16-DEC-2004 (first entry)
XX Novel S. pneumoniae DNA sequence, SEQ ID 1976.
XX Meningitis; bacteraemia; pneumonia; otitis media; ds;
KW bacterial infection.
XX Streptococcus pneumoniae.
OS

XX US6800744-B1.
XX 05-OCT-2004.
XX 30-JUN-1998; 98US-00107433.
XX 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Doucette-Stamm LA, Bush D;
XX WPI; 2004-697205/68.
DR P-PSDB; ADR95944.
XX New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX Disclosure; SEQ ID NO 1976; 151pp; English.
XX The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae nucleic acid sequences. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ Sequence 1692 BP; 567 A; 387 C; 332 G; 406 T; 0 U; 0 Other;
Query Match 32.3%; Score 85; DB 13; Length 1692;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
OY 90 TCCTGACCACCTATNTGTCATCAAGTGCACCAATGACACGTCGAGTGTGGGTTAGACAACT 149
DB 983 TGCTGACAACTTCTCTCCACCAAGTGCTTACGACCAAGACGAGTTATCTAGACAACT 1042
OY 150 ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTTAGCCAAATG 209
DB 1043 ACTATACCAACCAACCTATCATGATTGGCTTGTATAAGGCTCTGACTCCCAACCAAGATG 1102
OY 210 CCCAGCGGTACTTCAAGAGTACCAAGAACTCAAGGAGCGGTCAAGCACCTGA 263
DB 1103 CCCAACGCTATTTAAACGGTATCAGAACTCAAGAAAGCTGTCAATATTGA 1156
RESULT 11
AEA57211
ID AEA57211 standard; DNA; 1692 BP.
XX AC AEA57211;
XX 25-AUG-2005 (first entry)
DT

XX Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:1976.
DE bacterial infection; Streptococcus pneumoniae infection; antibacterial;
KW vaccine; gene; ds.
XX Streptococcus pneumoniae.
OS US2005136404-A1.
XX 23-JUN-2005.
XX 10-JUL-2003; 2003US-00617320.
XX 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX (DOUC/) DOUCETTE-STAMM L A.
PA (BUSH/) BUSH D.
XX Doucette-Stamm LA, Bush D;
PI WPI; 2005-477576/48.
XX P-PSDB; AEA59814.
DR
XX New isolated nucleic acid molecules and encoded polypeptides useful for
PT diagnosing, preventing or treating bacterial infections, particularly
PT Streptococcus pneumoniae infection.
XX Claim 1; SEQ ID NO 1976; 144pp; English.
XX The invention relates to an isolated nucleic acid molecule for detecting,
CC preventing or treating pathological conditions resulting from bacterial
CC infection. The isolated nucleic acid comprises: (a) any of the 2603
CC nucleotide sequences of AEA5236 to AEA57838; (b) a nucleotide sequence
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
CC sequence of at least 8 nucleotides in length, where the sequence is
CC hybridizable to a nucleic acid having any of the nucleotide sequences in
CC (a). Also described: (1) a recombinant expression vector comprising the
CC above nucleic acid operably linked to a transcription regulatory element;
CC (2) a cell comprising the recombinant expression vector; (3) producing an
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
CC consisting of at least 8 nucleotides of any of AEA5236 to AEA57838; (5)
CC treating a subject for S. pneumoniae infection; (6) a recombinant or its
CC substantially pure preparation of an S. pneumoniae polypeptide or its
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;
CC (7) a vaccine composition for preventing or treating an S. pneumoniae
CC infection, comprising an amount of the above nucleic acid or polypeptide;
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;
CC (9) a computer readable medium having recorded the nucleotide sequences
CC of AEA5236 to AEA57838; (10) a computer based system for identifying
CC fragments of the Streptococcus genome of commercial importance. The
CC composition and methods are useful for diagnosing, preventing or treating
CC bacterial infections, particularly S. pneumoniae infection. The present
CC sequence represents a S. pneumoniae ORF nucleic acid sequence from the
CC present invention. Note - The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from the USPTO web site.
XX Sequence 1692 BP; 567 A; 387 C; 332 G; 406 T; 0 U; 0 Other;
SQ Query Match 32.3%; Score 85; DB 14; Length 1692;
Best Local Similarity 67.8%; Pred. No. 3.6e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACCTATNTGCATCAAGTGCCTCAATGACCAAGTCAGTGTGGCTTAGACAAC 149
DB 983 TGCTGCAACCTTCCTCCACCAAGTGCTTAACGACCAAGACCGAGGTATCTTAGACAAC 1042
QY 150 ACTATACGGCAGGAACTGAGATTGAGTTGGATGTGGCTTTTGACTCTCTAGCCAAATG 209
|||||

DB 1043 ACTATACCAACCAACCTATCATGATTTGGCTTGATAGGCTTCTGACTCCCAACCAATG 1102
QY 210 CCCAGCGGTACTTCAAGAAAGTACCAAGAACTCAAGGAGCGGTCAAGCACTCTGA 263
|||||
DB 1103 CCCACGCTATTTTAAACGGTATCAGAACTCAAGAGAGCTGTCAATACTTGA 1156
|||||
RESULT 12
AAV52165/c
ID AAV52165 standard; DNA; 10240 BP.
XX AAV52165;
AC AAV52165;
XX 23-OCT-1998 (first entry)
DT Streptococcus pneumoniae genome fragment SEQ ID NO:32.
XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
DE Streptococcus pneumoniae; S. pneumoniae; vaccine; pharmaceutical composition; ds.
KW computer readable medium; vaccine; pharmaceutical composition; ds.
KW Streptococcus pneumoniae.
OS WO9818931-A2.
XX 07-MAY-1998.
XX 30-OCT-1997; 97WO-US019588.
PR 31-OCT-1996; 96US-0029960P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI Dougherty BA;
PI WPI; 1998-272225/24.
DR Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX Claim 1; Page 326-332; 1409pp; English.
XX The present invention describes a computer readable medium which has the
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC 391, identifying members of the library which contain sequences that
CC hybridize to the target sequence and isolating the nucleic acid molecules
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC homologous to amplification primers derived from the fragment of the S.
CC pneumoniae genome to prime the amplification and isolating the amplified
CC sequences. The computer readable medium can be used in a computer-based
CC system for identifying fragments of the S. pneumoniae genome of
CC commercial importance, or expression modulating fragments of the S.
CC pneumoniae genome. Products from the present invention can be used in
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC for S. pneumoniae
XX Sequence 10240 BP; 2865 A; 1914 C; 2390 G; 3068 T; 0 U; 3 Other;
SQ Query Match 32.3%; Score 85; DB 2; Length 10240;
Best Local Similarity 67.8%; Pred. No. 8.1e-18;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 90 TCCTGACCACCTATNTGCATCAAGTGCCTCAATGACCAAGTCAGTGTGGCTTAGACAAC 149

```

Db      4646  TGCTGACCACTTCTCCACCAAGTGCTTACGACCAAGACGAGTTATCTAGCAACT 4587
QY      150  ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTAGCCAAATG 209
Db      4586  ACTATACCAACCAACCTATCATGATTGGCTTGATAGGCTCTGACTCCCAACCAAGATG 4527
QY      210  CCCAGCGGTACTTCAAGAGTACCAAACTCAAGAGGCGGTCAAGCACTGA 263
Db      4526  CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAATACTTGA 4473

RESULT 13
ABS56454_07/c
Continuation (8 of 22) of ABS56454 from base 700001 (Streptococcus pneumoniae type 4 str
WP Sequence split into 22 fragments LOCUS ABS56454 Accession Abs56454
WP Fragment Name      Begin      End
WP ABS56454_00        1        110000
WP ABS56454_01       100001   210000
WP ABS56454_02       200001   310000
WP ABS56454_03       300001   410000
WP ABS56454_04       400001   510000
WP ABS56454_05       500001   610000
WP ABS56454_06       600001   710000
WP ABS56454_07       700001   810000
WP ABS56454_08       800001   910000
WP ABS56454_09       900001  1010000
WP ABS56454_10      1000001  1100000
WP ABS56454_11      1100001  1200000
WP ABS56454_12      1200001  1310000
WP ABS56454_13      1300001  1410000
WP ABS56454_14      1400001  1510000
WP ABS56454_15      1500001  1610000
WP ABS56454_16      1600001  1710000
WP ABS56454_17      1700001  1810000
WP ABS56454_18      1800001  1910000
WP ABS56454_19      1900001  2010000
WP ABS56454_20      2000001  2110000
WP ABS56454_21      2100001  2162598

Query Match      32.3%; Score 85; DB 10; Length 110000;
Best Local Similarity 67.8%; Pred. No. 2.3e-17;
Matches 118; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY      90  TCCTGACCACTTATNTGCATCAAGTGCACCAATGACCAATGACCAATGAGTGTGGTTAGCAACT 149
Db      7483  TGCTGACCACTTCTCCACCAAGTGCTTACGACCAAGACCAAGTATCTAGCAACT 7424

QY      150  ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTAGCCAAATG 209
Db      7423  ACTATACCAACCAACCTATCATGATTGGCTTGATAGGCTCTGACTCCCAACCAAGATG 7364

QY      210  CCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGAGGCGGTCAAGCACTGA 263
Db      7363  CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAATACTTGA 7310

RESULT 15
ACA50476
ID ACA50476 standard; DNA; 1653 BP.
XX
AC ACA50476;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #32133.
XX
KW Antisense; db; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Streptococcus pyogenes.
XX
FN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-FSDB; ABU46606.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 38346; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
```

```

Db      4646  TGCTGACCACTTCTCCACCAAGTGCTTACGACCAAGACGAGTTATCTAGCAACT 4587
QY      150  ACTATACGGGCAAGAACTGGAGATTGAGTTGGATTGGCTTTGACTCTAGCCAAATG 209
Db      4586  ACTATACCAACCAACCTATCATGATTGGCTTGATAGGCTCTGACTCCCAACCAAGATG 4527
QY      210  CCCAGCGGTACTTCAAGAGTACCAAACTCAAGAGGCGGTCAAGCACTGA 263
Db      4526  CCCAACGCTATTTTAAACGGTATCAGAACTCAAGAAGCTGTCAAATACTTGA 4473

RESULT 14
ABS56454_08/c
Continuation (9 of 22) of ABS56454 from base 800001 (Streptococcus pneumoniae type 4 str
WP Sequence split into 22 fragments LOCUS ABS56454 Accession Abs56454
WP Fragment Name      Begin      End
WP ABS56454_00        1        110000
WP ABS56454_01       100001   210000
WP ABS56454_02       200001   310000
WP ABS56454_03       300001   410000
WP ABS56454_04       400001   510000
WP ABS56454_05       500001   610000
WP ABS56454_06       600001   710000
WP ABS56454_07       700001   810000
WP ABS56454_08       800001   910000
WP ABS56454_09       900001  1010000
WP ABS56454_10      1000001  1110000
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2006, 14:03:25 ; Search time 3837 seconds
(without alignments)

3206.933 Million cell updates/sec

Title: US-10-632-117-37

Perfect score: 263

Sequence: 1 acgaataatgatgatccat.....ggaggcggtcaagcacctga 263

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	14.4	428	5	BY453989
2	37.6	14.3	807	6	CB898978
3	37.4	14.2	1101	10	CN300L12
4	37	14.1	850	5	BX368104
5	36.8	14.0	406	7	CV072680
6	36.8	14.0	648	5	CA036082
7	36.2	13.8	427	3	BM638966
8	36.2	13.8	476	3	BM642048
9	36.2	13.8	486	3	BM636853
10	36.2	13.8	578	3	BM616582
11	36.2	13.8	585	3	BM624215
12	36.2	13.8	607	3	BM628044
13	36.2	13.8	610	3	BM625696
14	36.2	13.8	615	3	BM578413
15	36.2	13.8	630	3	BM618626
16	36.2	13.8	645	3	BM642729
17	36.2	13.8	648	3	BM627773
18	36.2	13.8	649	3	BM649206
19	36.2	13.8	661	3	BM648844
20	36.2	13.8	670	3	BM632022
21	36.2	13.8	674	3	BM636042
22	36.2	13.8	678	3	BM636340

23	36.2	13.8	683	3	BM651776
24	36.2	13.8	684	3	BM625442
25	36.2	13.8	691	3	BM643948
26	36.2	13.8	692	3	BM653819
27	36.2	13.8	693	3	BM656762
28	36.2	13.8	697	3	BM631562
29	36.2	13.8	709	3	BM592722
30	36.2	13.8	711	3	BM641312
31	36.2	13.8	714	3	BM646862
32	36.2	13.8	715	3	BM627009
33	36	13.7	307	10	CG618897
34	36	13.7	629	5	BM648199
35	35.8	13.6	614	3	BM686992
36	35.6	13.5	441	9	AZ108235
37	35.6	13.5	653	6	CA449674
38	35.6	13.5	713	2	BE970931
39	35.4	13.5	177	1	AW553179
40	35.4	13.5	289	2	BG146742
41	35.4	13.5	299	1	AA086663
42	35.4	13.5	323	10	CG887081
43	35.4	13.5	365	2	BE197026
44	35.4	13.5	500	1	AA689925
45	35.4	13.5	545	6	CA322313

ALIGNMENTS

RESULT 1

BY453989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

REFERENCE

AUTHORS

BY453989 RIKEN full-length enriched, pooled tissues, adult spleen, etc. Mus musculus cDNA clone K630070L23 3', mRNA sequence. linear EST 13-DEC-2002

BY453989.1 GI:26750723

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 428)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Kikado, I., Oatso, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusnic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziarski, R. M., King, B. I., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Negashima, T., Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tonita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakakuma, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Azawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation

Nature 420, 563-573 (2002)

Pubmed 12466851

COMMENT

Contact: Yoshihide Hayaahizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Aizawa.K., Akimura.T., Arakawa.T., Carninci.P., Fukuda.S., Hirozane.T., Imotani.K., Ishii.Y., Itoh.M., Kawai.J., Konno.H., Miyazaki.A., Murata.M., Nakamura.M., Nomura.K., Numazaki.R., Ohno.M., Sakai.K., Sakazume.N., Sasaki.D., Sato.K., Shibata.K., Shiraki.T., Tagami.M., Waki.K., Watahiki.A., Muramatsu.M. and Hayashizaki.Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

```

1..428
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, pooled tissues, adult spleen, etc."
/notes="(dev stage=adult, tissue type=spleen, sex=male), (dev stage=adult, tissue type=kidney, sex=male), (dev stage=adult, tissue type=testis, sex=male), (dev stage=adult, tissue type=thymus, sex=male), (dev stage=adult, tissue type=heart, sex=male), (dev stage=adult, tissue type=colon, sex=male), (dev stage=adult, tissue type=stomach, sex=male), (dev stage=adult, tissue type=liver, sex=male), (dev stage=13 days embryo, tissue type=whole body, sex=mix), (dev stage=14 days embryo, tissue type=whole body, sex=mix), (dev stage=16 days embryo, tissue type=whole body, sex=mix), (dev stage=17 days embryo, tissue type=whole body, sex=mix), (dev stage=15 days pregnant, adult, tissue type=amion, sex=female), (dev stage=10 days neonate, tissue type=brain, sex=mix), (dev stage=10 days neonate, tissue type=thymus, sex=mix), (dev stage=10 days neonate, tissue type=heart, sex=mix)"

```

ORIGIN

```

Query Match      14.4%; Score 38; DB 5; Length 428;
Best Local Similarity 54.8%; Pred. No. 0.87;
Matches 74; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 87 CCCTCTGACCACTATATGTCATCAAGTGCACAAATGACCAAGTCGAGTGTGGGTTAGACA 146
    |||||
Db 255 CCGCGGGCCAGCTGTAGCTGTATTGCTCGGACCCGCGCGCTGTGGCCGAGCG 314
    |||||
QY 147 ACTACTATACGGGCAAGGAACCTGGAGATTGAGTTGACTTGGCTTGTACTCTAGCCAAA 206
    |||||
Db 315 ACGGCTATGTGTCGAAGCCAGGGGCTGTGATTGTATCTGCGAAGATCCAAGCCCCGA 374
    |||||
QY 207 ATGCCACGGGTACT 221
    |||||

```

Db

375 AAGCCATGAGCCT 389

RESULT 2

CB898978

LOCUS

DEFINITION

CB898978 807 bp mRNA linear EST 02-JUL-2003

ACCESSION

CB898978

VERSION

EST.

KEYWORDS

EST.

SOURCE

ORGANISM

Hypocrea jecorina (anamorph: Trichoderma reesei)

Eukaryota; Fungi; Ascomycota; Peizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.

1 (bases 1 to 807)

Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S., Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J., Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C., Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.

Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus Trichoderma reesei

J. Biol. Chem. 278 (34), 31988-31997 (2003)

Contact: Pamela K. Foreman

12788920

Genecor Intl.

925 Page Mill Road, Palo Alto, CA 94304, USA

Tel: (650) 846-7635

Fax: (650) 621-7817

Email: Pforeman@genecor.com

Seq primer: LT-F1 primer.

Location/Qualifiers

1..807

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/strain="QM6a"

/db_xref="taxon:51453"

/clone="tric016xml0"

/dev_stage="mycelia"

/clone_lib="T.reesei mycelial culture, Version 3 april"

/note="Vector: pREP3Y; Site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match

Best Local Similarity

Matches

80; Conservative

0; Mismatches

116; Indels

0; Gaps

0;

QY 61 TTTTGTGATCGAATTCGAGCTGCCCTCTGACCACTATATGTCATCAAGTGCACAA 120

|||||

Db 1 TTTTGTGNN 60

QY 121 TGACCAAGTCGAGTGTGGGTTAGACAACACTACTATACGGGCAAGGAACCTGGAGATTGAGTT 180

Db 61 CGTCGGCAGATTCTCAAGGAGCGCAAGCTCAAGAGAGCGAGGTTGACGACATCTTCT 120

QY 181 GGATGTGGCTTTGACTCTCTAGCCAAAATGCCAGCGGTPACTTCAAGAGTACCAAGAACT 240

Db 121 GGTGGGGGTTTCCACCCCGTATCCCAAGGTTTCAGTCTCTTATCGAGGAGTACTTTAAGCG 180

QY 241 CAAGAGGCGGTCAAG 256

Db 181 CAAGAAGGCTTCCAAG 196

RESULT 3

CN800112

LOCUS

DEFINITION

CN800112/c

Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR23D10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

1101 bp DNA linear GSS 03-JUN-1999

AL067834

ACCESSION

```

VERSION AL067834.1 GI:4957863
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
        Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
        Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
        Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPci-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
    source
        Location/Qualifiers
            1..1101
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /db_xref="taxon:7227"
                /clone="BACR23D10"
                /clone_lib="RPci-98"
                /note="end : TET3"
ORIGIN
Query Match 14.2%; Score 37.4; DB 10; Length 1101;
Best Local Similarity 33.2%; Pred. No. 1.7;
Matches 68; Conservative 47; Mismatches 90; Indels 0; Gaps 0;
QY 58 TATTTTTCGTGAATCGAATTCGAGCTCGCCCTCTCGACACCTATNTGCATCAAGTGCC 117
Db 1051 TATTTTTCGTGAATTCGAGCTCGCCCTCTCGACACCTATNTGCATCAAGTGCC 117
QY 118 AAATGACACGAGTGCGGTGTAGACAACTACTATACGGCAAGGAACTGGAGATTCA 177
Db 991 AAAAAAAAAWGGRRGGVADDDGRAMDMAVCCHAAMSVVAAAMTMCCAAAGVSGSGR 932
QY 178 GTTGGATGTGGTTTCACTCTAGCCAAATGCCCAGCGGTACTTCAAGAAGTACAGAA 237
Db 931 DNCGGGGGGAAGGSDSTCCCBKCAACACGCGSVGAGRWDDGGRRRAAATWDAAGG 872
QY 238 ACTCAGAGGCGGTCAAGCACTG 262
Db 871 GGGRAAAARGGGRRDRAWAAACKG 847
RESULT 4
BX368104/c
LOCUS BX368104 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1073YC17 3-PRIME, mRNA sequence.
ACCESSION BX368104
VERSION BX368104
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
        Homnidae; Homo.
REFERENCE 1 (bases 1 to 850)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gascon Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
7464.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?e=CS0AU013ZD11_U01236_l&c=7464.f.
FEATURES
    source
        Location/Qualifiers
            1..850
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0D1073YC17"
                /clone_lib="PLACENTA COT 25-NORMALIZED"
                /clone_type="Homo sapiens PLACENTA COT 25-NORMALIZED"
                /note="1st strand cDNA was primed with a NotI-oligo(dT)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and EcoR V sites of the pCMVSPORT 6
                vector. Library was normalized."
ORIGIN
Query Match 14.1%; Score 37; DB 5; Length 850;
Best Local Similarity 54.3%; Pred. No. 2.2;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 135 TCGCGTTAGACAACTACTATACGGCAAGGAACTGAGATTGAGTTGGCTTGA 194
Db 775 TCGCGTTCAGCACTACTATACGGCAAGGAACTGAGATTGAGTTGGCTTGA 194
QY 195 CTCCTAGCCAAATGCCAGCGGTACTTCAAGAAGTACCAGAACTCAAGGAGCGGTCA 254
Db 715 CTCGTATCCCCAAGATCCAGAGCTGCTCAGAGATTCTTCAACGCGCAAGGAGCTCA 656
QY 255 AGCAGCTGA 263
Db 655 AGAGCATCA 647
RESULT 5
CV072680
LOCUS Lr JV2CF 27G04 SKplus Juvenile Earthworm Library Lumbricus rubellus
DEFINITION cDNA clone Lr JV2CF_27G04, mRNA sequence.
ACCESSION CV072680
VERSION CV072680
KEYWORDS EST.
SOURCE Lumbricus rubellus (humus earthworm)
ORGANISM Lumbricus rubellus
        Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
        Lumbricina; Lumbricidae; Lumbricus.
REFERENCE 1 (bases 1 to 406)
AUTHORS Chaiseley, J., Hedley, B.A., Morgan, A.J., Sturzenbaum, S., Kille, P. and
        Blaxter, M.
TITLE The Lumbricus rubellus EST program - Sequences from a Juvenile
JOURNAL Unpublished (2004)
COMMENT Contact: Jennifer Chaiseley
BIOS1 1
Cardiff University
Main College, Museum Avenue, Cardiff, CF11 3TL, UK
Tel: +44 2920876680
Fax: +44 2920874305
Email: chaiseley@cardiff.ac.uk, Kille@cardiff.ac.uk
Sequencing was performed in Edinburgh using the pBluescript II XR
cDNA library (Stratagene) protocol.

```

```

PCR Primers
FORWARD: M13P (gtttccagtcacagcg)
BACKWARD: M13R (caggaacagctatgaccatg)
Plate: 27 row: G column: 04
Seq primer: SKplus
High quality sequence start: 8
High quality sequence stop: 351.
FEATURES
    source
        location/Qualifiers
            1..406
                /organism="Lumbricus rubellus"
                /mol_type="mRNA"
                /db_xref="taxon:35632"
                /clone="Lr_JV2CF_27G04"
                /sex="mixed"
                /tissue_type="Whole worm"
                /dev_stage="Juvenile"
                /clone_lib="Juvenile Earthworm Library"
                /note="Organ: Whole worm"
ORIGIN
Query Match          14.0%; Score 36.8; DB 7; Length 406;
Best Local Similarity 66.2%; Pred. No. 2.1;
Matches 53; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 160 CAAGAACTGGAGATTGAGTTGGATGGCTTTGACTCTAGCCAAAATGCCAGCGTA 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97 CAACGAATGGAGTTGCTTGGATGGAGCGAACAGAGACGCCGAGAGCGCGAGAAGAA 156
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 220 CTTCAAGAAGTACCAGAAAC 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 157 CATCAGAAGTTCACGCAAC 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 6
CA036082
LOCUS
DEFINITION
    Lr_ade_01G12_T3 Earthworm Lambda Zap Express Library Lumbricus
    rubellus cDNA clone Lr_ade_01G12 5' similar to pir|A59287 myosin
    heavy chain - fluke (Schistosoma mansoni) (strain Brazilian, mRNA
    sequence.
ACCESSION
    CA036082
VERSION
    CA036082.1 GI:24334779
KEYWORDS
    EST.
SOURCE
    Lumbricus rubellus (humus earthworm)
ORGANISM
    Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
    Lumbricina; Lumbricidae; Lumbricidae.
REFERENCE
    1 (bases 1 to 648)
AUTHORS
    Sturzenbaum,S., Parkinson,J., Blaxter,M., Morgan,J., Kille,P.,
    Schaffner,W. and Georgiev,O.
TITLE
    Expressed Sequence Tags from the humus earthworm L. rubellus
JOURNAL
    Unpublished (2000)
COMMENT
    Contact: Blaxter ML
    Institute of Cell, Animal and Population Biology
    University of Edinburgh
    Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
    3JT, UK.
    Tel: +44 131 650 6760
    Fax: +44 131 670 5450
    Email: mark.blaxter@ed.ac.uk
    The library was prepared using protocol given by supplier
    (Stratagene).
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 01 row: G column: 12
Seq primer: T3
High quality sequence stop: 457.
FEATURES
    source
        location/Qualifiers
            1..648
                /organism="Lumbricus rubellus"
                /mol_type="mRNA"
                /db_xref="taxon:35632"

```

```

/clone="Lr_ade_01G12"
/tissue_type="Whole worm"
/dev_stage="adult"
/clone_lib="Earthworm Lambda Zap Express Library"
/note="Vector: pBK-CMV; The library was prepared using
protocol given by supplier (Stratagene)."
ORIGIN
Query Match          14.0%; Score 36.8; DB 5; Length 648;
Best Local Similarity 66.2%; Pred. No. 2.4;
Matches 53; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 160 CAAGAACTGGAGATTGAGTTGGATGGCTTTGACTCTAGCCAAAATGCCAGCGTA 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 275 CAACGAATGGAGTTGCTTGGATGGAGCGAACAGAGACGCCGAGAGCGCGAGAAGAA 334
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 220 CTTCAAGAAGTACCAGAAAC 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 335 CATCAGAAGTTCACGCAAC 354
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 7
BM638966
LOCUS
DEFINITION
    17000687567125 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
    19600449621565 5', mRNA sequence.
ACCESSION
    BM638966
VERSION
    BM638966.1 GI:18938477
KEYWORDS
    EST.
SOURCE
    Anopheles gambiae (African malaria mosquito)
ORGANISM
    Anopheles gambiae
    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
    Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
    Culicidae; Anophelinae; Anopheles.
REFERENCE
    1 (bases 1 to 427)
AUTHORS
    Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
    Charlab,R., Collins,P.H., Venter,J.C. and Hoffman,S.L.
TITLE
    Celera Anopheles gambiae EST project
JOURNAL
    Unpublished (2002)
COMMENT
    Contact: Holt R.A.
    Celera Genomics
    45 W. Gude Dr., Rockville, MD 20850, USA
    Tel: 2404533151
    Fax: 2404534580
    Email: HoltRA@celera.com
    Plate: NU01004N8J row: E column: 19
    Seq primer: M13 Reverse.
FEATURES
    source
        location/Qualifiers
            1..427
                /organism="Anopheles gambiae"
                /mol_type="mRNA"
                /strain="RSP-ST (Reduced susc. to Permethrin - std.
                chromosome)"
                /db_xref="taxon:7165"
                /clone="19600449621565"
                /dev_stage="Adult"
                /lab_host="DHI0b"
                /clone_lib="A.Gam.ad.cDNA1"
                /note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
                adult mosquitoes (mixed sex)-frozen on liquid nitrogen.
                cDNA inserts >500 bp cloned directionally into pSport 1.
                Not 1 site is 3'. Clones available through the Malaria
                Research and Reference Reagent Resource Center
                (www.malaria.mr4.org)."
ORIGIN
Query Match          13.8%; Score 36.2; DB 3; Length 427;
Best Local Similarity 60.8%; Pred. No. 3.3;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 160 CAAGAACTGGAGATTGAGTTGGATGGCTTTGACTCTAGCCAAAATGCCAGCGTA 219
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 211 CAACGAGCTGGAGATTGCTCTGGATCAGGCCAACAGAGCTAACGCTGAGGCCGAGAAGAA 270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY      220  CTTCAAGAGTACCAAGAACTCAAGGCGGTCAAG 256
DB      271  CATCAAGCGCTACCAGCAGCAGCTGAAGGACGTCCAG 307

RESULT 8
LOCUS   BM642048
DEFINITION 17000687309375 A.Gam.ad.cdNAL Anopheles gambiae cDNA clone
VERSION 19600449651236 5', mRNA sequence.
ACCESSION BM642048
KEYWORDS 19600449651236 5', mRNA sequence.
SOURCE  BM642048.1 GI:18941559
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 476)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celerla Anopheles gambiae EST project
Unpublished (2002)
CONTACT: Holt R.A.
Celerla Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU010049W6 row: J column: 02
Seq primer: M13 Reverse
Location/Qualifiers
FEATURES
source
1..476
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449651236"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNAL"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 476;
Best Local Similarity 60.8%; Pred. No. 3.4;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY      160  CAAGGAAGTGGAGTGGATGGATGGCTTGTGACTCTAGCCAAATGCCAGCGTA 219
DB      249  CAACGAGCTGGAGTGTCTGTGATCAGCCAAAGGCTAACGCTGAGGCCGAGAA 308

RESULT 9
LOCUS   BM636853
DEFINITION 17000687562979 A.Gam.ad.cdNAL Anopheles gambiae cDNA clone
VERSION 19600449642127 5', mRNA sequence.
ACCESSION BM636853
KEYWORDS 19600449642127 5', mRNA sequence.
SOURCE  BM636853.1 GI:18936364
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 486)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celerla Anopheles gambiae EST project
Unpublished (2002)
CONTACT: Holt R.A.
Celerla Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01003CV2 row: N column: 13
Seq primer: M13 Reverse
Location/Qualifiers
FEATURES
source
1..486
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449642127"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNAL"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 486;
Best Local Similarity 60.8%; Pred. No. 3.4;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY      160  CAAGGAAGTGGAGTGGATGGATGGCTTGTGACTCTAGCCAAATGCCAGCGTA 219
DB      277  CAACGAGCTGGAGTGTCTGTGATCAGCCAAAGGCTAACGCTGAGGCCGAGAA 336

RESULT 10
LOCUS   BM616582
DEFINITION 17000687148935 A.Gam.ad.cdNAL bloodi Anopheles gambiae cDNA clone
VERSION 19600449713643 5', mRNA sequence.
ACCESSION BM616582
KEYWORDS 19600449713643 5', mRNA sequence.
SOURCE  BM616582.1 GI:18914804
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 578)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celerla Anopheles gambiae EST project
Unpublished (2002)
CONTACT: Holt R.A.
Celerla Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com

```

```

ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 486)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celerla Anopheles gambiae EST project
Unpublished (2002)
CONTACT: Holt R.A.
Celerla Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01003CV2 row: N column: 13
Seq primer: M13 Reverse
Location/Qualifiers
FEATURES
source
1..486
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449642127"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdNAL"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 486;
Best Local Similarity 60.8%; Pred. No. 3.4;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY      160  CAAGGAAGTGGAGTGGATGGATGGCTTGTGACTCTAGCCAAATGCCAGCGTA 219
DB      277  CAACGAGCTGGAGTGTCTGTGATCAGCCAAAGGCTAACGCTGAGGCCGAGAA 336

RESULT 10
LOCUS   BM616582
DEFINITION 17000687148935 A.Gam.ad.cdNAL bloodi Anopheles gambiae cDNA clone
VERSION 19600449713643 5', mRNA sequence.
ACCESSION BM616582
KEYWORDS 19600449713643 5', mRNA sequence.
SOURCE  BM616582.1 GI:18914804
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 578)
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celerla Anopheles gambiae EST project
Unpublished (2002)
CONTACT: Holt R.A.
Celerla Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com

```

```

Plate: NU01004YX row: A column: 21
Seq primer: M13 Reverse.
Location/Qualifiers
1. .578
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449713643"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdna.blood1"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 578;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGAACTGGAGATTGATGGCTTTGACTCTCTAGCCAAATGCCAGCGGTA 219
Db 38 CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACAGGCTAACGCTGAGGCCCAAGAA 97

QY 220 CTTCAAGAGTACCAAGAACTCAAGGCGGTCAG 256
Db 98 CATCAAGCGCTACCAGACGACTGAAGGACGTCACAG 134

RESULT 11
LOCUS BM624215 585 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687490727 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
ACCESSION BM624215
VERSION 1960044969056 5', mRNA sequence.
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 585)
AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,
Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
JOURNAL
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004YV row: P column: 14
Seq primer: M13 Reverse.
Location/Qualifiers
1. .585
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="1960044969056"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdna1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.

FEATURES
source

```

```

cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 585;
Best Local Similarity 60.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGAACTGGAGATTGATGGCTTTGACTCTCTAGCCAAATGCCAGCGGTA 219
Db 466 CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACAGGCTAACGCTGAGGCCCAAGAA 525

QY 220 CTTCAAGAGTACCAAGAACTCAAGGCGGTCAG 256
Db 526 CATCAAGCGCTACCAGACGACTGAAGGACGTCAG 562

RESULT 12
LOCUS BM628044 607 bp mRNA linear EST 26-FEB-2002
DEFINITION 17000687497740 A.Gam.ad.cdna1 Anopheles gambiae cDNA clone
ACCESSION BM628044
VERSION 19600449652861 5', mRNA sequence.
KEYWORDS EST.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Culicidae; Anophelinae; Anopheles.
REFERENCE 1 (bases 1 to 607)
AUTHORS Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,
Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
JOURNAL
COMMENT Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: HoltRA@celera.com
Plate: NU01004YV row: M column: 19
Seq primer: M13 Reverse.
Location/Qualifiers
1. .607
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449652861"
/dev_stage="Adult"
/lab_host="DH10b"
/clone_lib="A.Gam.ad.cdna1"
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN
Query Match 13.8%; Score 36.2; DB 3; Length 607;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 160 CAAGAACTGGAGATTGATGGCTTTGACTCTCTAGCCAAATGCCAGCGGTA 219
Db 170 CAACGAGCTGGAGATTGCTCTGGATCAGCCCAACAGGCTAACGCTGAGGCCCAAGAA 229

QY 220 CTTCAAGAGTACCAAGAACTCAAGGCGGTCAG 256

```


[illegible]

